SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT: Altieri, Dario C.

    (ii) TITLE OF INVENTION: SURVIVIN, A PROTEIN THAT INHIBITS
        CELLULAR APOPTOSIS, AND ITS MODULATION

    (iii) NUMBER OF SEQUENCES: 35

    (iv) CORRESPONDENCE ADDRESS:
        (A) ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
        (B) STREET: 1800 M Street, N.W.
        (C) CITY: Washington
        (D) STATE: D.C.
        (E) COUNTRY: USA
        (F) ZIP: 20036-5869

    (v) COMPUTER READABLE FORM:
        (A) MEDIUM TYPE: Floppy disk
        (B) COMPUTER: IBM PC compatible
        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
        (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

    (vi) CURRENT APPLICATION DATA:
        (A) APPLICATION NUMBER: US 08/975,080
        (B) FILING DATE: 20-NOV-1997

    (vii) PRIOR APPLICATION DATA:
        (A) APPLICATION NUMBER: US 60/031,435
        (B) FILING DATE: 20-NOV-1996

    (viii) ATTORNEY/AGENT INFORMATION:
        (A) NAME: Adler, Reid G.
        (B) REGISTRATION NUMBER: 30,988
        (C) REFERENCE/DOCKET NUMBER: 044574-5022-01-WO

    (ix) TELECOMMUNICATION INFORMATION:
        (A) TELEPHONE: 202-467-7000
        (B) TELEFAX: 202-467-7176

(2) INFORMATION FOR SEQ ID NO:1:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 19 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: other nucleic acid
        (A) DESCRIPTION: /desc = "oligonucleotide"

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGCTGGCCGC TCCTCCCTC . 19

(2) INFORMATION FOR SEQ ID NO:2:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 18 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: other nucleic acid
        (A) DESCRIPTION: /desc = "oligonucleotide"

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGACCTCCA GAGGTTTC 18

(2) INFORMATION FOR SEQ ID NO:3:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 17 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ala Pro Thr Leu Pro Pro Ala Trp Gln Pro Phe Leu Lys Asp His Arg
1   5   10   15

Ile

(2) INFORMATION FOR SEQ ID NO:4:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 20 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Glu Gly Trp Glu Pro Asp Asp Asp Pro Ile Glu Glu His Lys Lys His
1   5   10   15

Ser Ser Gly Cys
20

(2) INFORMATION FOR SEQ ID NO:5:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 27 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCGGGTGAGC TGTCCCTTGC AGATGGC   27

(2) INFORMATION FOR SEQ ID NO:6:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 27 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCATGTAAGT TGATTTTTCT AGAGAGG   27

(2) INFORMATION FOR SEQ ID NO:7:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 27 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AATTGTATGT CTTTATTTCC AGGCAAA   27

(2) INFORMATION FOR SEQ ID NO:8:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Glu Glu Ala Arg Leu Val Thr Phe Gln Asn Trp Pro Asp Ala Phe Leu
1   5   10   15

Thr Pro Gln Glu Leu Ala Lys Ala Gly Phe Tyr Tyr Leu Gly Arg Gly
            20   25   30

Asp Gln Val Gln Cys Phe Ala Cys Gly Gly Lys Leu Ala
        35   40   45

(2) INFORMATION FOR SEQ ID NO:9:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Glu Glu Ala Arg Phe Leu Thr Tyr Ser Met Trp Pro Leu Ser Phe Leu
1   5   10   15

Ser Pro Ala Glu Leu Ala Arg Ala Gly Phe Tyr Tyr Ile Gly Pro Gly
            20   25   30

Asp Arg Val Ala Cys Phe Ala Cys Gly Gly Lys Leu Ser
        35   40   45

(2) INFORMATION FOR SEQ ID NO:10:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Glu Ala Asn Arg Leu Val Thr Phe Lys Asp Trp Pro Asn Pro Asn Ile
1   5   10   15

Thr Pro Gln Ala Leu Ala Lys Ala Gly Phe Tyr Tyr Leu Asn Arg Leu
            20   25   30

Asp His Val Lys Cys Val Trp Cys Asn Gly Val Ile Ala
        35   40   45

(2) INFORMATION FOR SEQ ID NO:11:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Glu Glu Val Arg Leu Asn Thr Phe Glu Lys Trp Pro Val Ser Phe Leu
1               5                   10                  15

Ser Pro Glu Thr Met Ala Lys Asn Gly Phe Tyr Tyr Leu Gly Arg Ser
            20              25                  30

Asp Glu Val Arg Cys Ala Phe Cys Lys Val Glu Ile Met
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO:12:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Lys Ala Ala Arg Leu Gly Thr Tyr Thr Asn Trp Pro Val Gln Phe Leu
1               5                   10                  15

Glu Pro Ser Arg Met Ala Ala Ser Gly Phe Tyr Tyr Leu Gly Arg Gly
            20              25                  30

Asp Glu Val Arg Cys Ala Phe Cys Lys Val Glu Ile Thr
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO:13:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 47 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Glu Glu Ala Arg Leu Ala Ser Phe Arg Asn Trp Pro Phe Tyr Val Gln
1               5                   10                  15

Gly Ile Ser Pro Cys Val Leu Ser Glu Ala Gly Phe Val Phe Thr Gly
            20                  25                  30

Lys Gln Asp Thr Val Gln Cys Phe Ser Cys Gly Gly Cys Leu Gly
            35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO:14:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Glu Ala Asn Arg Leu Val Thr Phe Lys Asp Trp Pro Asn Pro Asn Ile
1               5                   10                  15

Thr Pro Gln Ala Leu Ala Lys Ala Gly Phe Tyr Tyr Leu Asn Arg Leu
            20                  25                  30

Asp His Val Lys Cys Val Trp Cys Asn Gly Val Ile Ala
            35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO:15:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 46 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
Glu Glu Ala Arg Leu Lys Ser Phe Gln Asn Trp Pro Asp Tyr Ala His
1               5                   10                  15

Leu Thr Pro Arg Glu Leu Ala Ser Ala Gly Leu Tyr Tyr Thr Gly Ile
            20                  25                  30

Gly Asp Gln Val Gln Cys Phe Cys Cys Gly Gly Lys Leu Lys
            35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO:16:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 46 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Glu Glu Ala Arg Leu Lys Ser Phe Gln Asn Trp Pro Asp Tyr Ala His
1               5                   10              15

Leu Thr Pro Arg Glu Leu Ala Ser Ala Gly Leu Tyr Tyr Thr Gly Ala
            20              25              30

Asp Asp Gln Val Gln Cys Phe Cys Cys Gly Gly Lys Leu Glu
        35               40              45
```

(2) INFORMATION FOR SEQ ID NO:17:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
Glu Asn Ala Arg Leu Leu Thr Phe Gln Thr Trp Pro Leu Thr Phe Leu
1               5                   10              15

Ser Pro Thr Asp Leu Ala Arg Ala Gly Phe Tyr Tyr Thr Gly Pro Gly
            20              25              30

Asp Arg Val Ala Cys Phe Ala Cys Gly Gly Lys Leu Ser
        35              40              45
```

(2) INFORMATION FOR SEQ ID NO:18:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Glu Glu Ala Arg Phe Leu Thr Tyr His Met Trp Pro Leu Thr Phe Leu
1               5               10              15

Ser Pro Ser Glu Leu Ala Arg Ala Gly Phe Tyr Tyr Ile Gly Pro Gly
            20              25              30

Asp Arg Val Ala Cys Phe Ala Cys Gly Gly Lys Leu Ser
        35              40              45
```

(2) INFORMATION FOR SEQ ID NO:19:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 46 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
Glu Glu Ala Arg Leu Lys Ser Phe Gln Asn Trp Pro Asp Tyr Ala His
1               5               10              15

Leu Thr Pro Arg Glu Leu Ala Ser Ala Gly Leu Tyr Tyr Thr Gly Ile
            20              25              30

Gly Asp Gln Val Gln Cys Phe Cys Cys Gly Gly Lys Leu Lys
        35              40              45
```

(2) INFORMATION FOR SEQ ID NO:20:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Glu Ala Asn Arg Leu Val Thr Phe Lys Asp Trp Pro Asn Pro Asn Ile
1               5               10              15

Thr Pro Gln Ala Leu Ala Lys Ala Gly Phe Tyr Tyr Leu Asn Arg Leu
            20              25              30

Asp His Val Lys Cys Val Trp Cys Asn Gly Val Ile Ala
        35              40              45
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 50 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS:
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
Tyr Val Gly Ile Gly Asp Lys Val Lys Cys Phe His Cys Asp Gly Gly
1               5                   10                  15

Leu Arg Asp Trp Glu Pro Gly Asp Asp Pro Trp Glu Glu His Ala Lys
            20                  25                  30

Trp Phe Pro Arg Cys Glu Phe Leu Leu Leu Ala Lys Gly Gln Glu Tyr
        35                  40                  45

Val Ser
    50
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 50 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS:
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Tyr Val Asp Arg Asn Asp Asp Val Lys Cys Phe Cys Cys Asp Gly Gly
1               5                   10                  ·15

Leu Arg Cys Trp Glu Pro Gly Asp Asp Pro Trp Ile Glu His Ala Lys
            20                  25                  30

Trp Phe Pro Arg Cys Glu Phe Leu Ile Arg Met Lys Gly Gln Glu Phe
        35                  40                  45

Val Asp
    50
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 50 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS:
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
Tyr Gln Lys Ile Gly Asp Gln Val Arg Cys Phe His Cys Asn Ile Gly
1               5                   10                  15

Leu Arg Ser Trp Gln Lys Glu Asp Glu Pro Trp Phe Glu His Ala Lys
            20                  25                  30

Trp Ser Pro Lys Cys Gln Phe Val Leu Leu Ala Lys Gly Pro Ala Tyr
        35                  40                  45

Val Ser
    50
```

(2) INFORMATION FOR SEQ ID NO:24:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 49 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Tyr Thr Gly Tyr Gly Asp Asn Thr Lys Cys Phe Tyr Cys Asp Gly Gly
1               5                   10                  15

Leu Lys Asp Trp Glu Pro Glu Asp Val Pro Trp Glu Gln His Val Arg
            20                  25                  30

Trp Phe Asp Arg Cys Ala Tyr Val Gln Leu Val Lys Gly Arg Asp Tyr
        35                  40                  45

Val
```

(2) INFORMATION FOR SEQ ID NO:25:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 49 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
Tyr Thr Gly Gln Gly Asp Lys Thr Arg Cys Phe Cys Cys Asp Gly Gly
1               5                   10                  15
```

```
Leu Lys Asp Trp Glu Pro Asp Asp Ala Pro Trp Gln Gln His Ala Arg
         20                  25                  30
```

```
Trp Tyr Asp Arg Cys Glu Tyr Val Leu Leu Val Lys Gly Arg Asp Phe
         35                  40                  45
```

```
Val
```

(2) INFORMATION FOR SEQ ID NO:26:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 50 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Tyr Thr Gly Ile Lys Asp Ile Val Gln Cys Phe Ser Cys Gly Gly Cys
1                5                  10                  15
```

```
Leu Glu Lys Trp Gln Glu Gly Asp Asp Pro Leu Asp Asp His Thr Arg
         20                  25                  30
```

```
Cys Phe Pro Asn Cys Pro Phe Leu Gln Asn Met Lys Ser Ser Ala Glu
         35                  40                  45
```

```
Val Thr
50
```

(2) INFORMATION FOR SEQ ID NO:27:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 50 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
Tyr Gln Lys Ile Gly Asp Gln Val Arg Cys Phe His Cys Asn Ile Gly
1                5                  10                  15
```

```
Leu Arg Ser Trp Gln Lys Glu Asp Glu Pro Trp Phe Glu His Ala Lys
         20                  25                  30
```

```
Trp Ser Pro Lys Cys Gln Phe Val Leu Leu Ala Lys Gly Pro Ser Tyr
         35                  40                  45
```

Val Ser
            50

(2) INFORMATION FOR SEQ ID NO:28:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 50 amino acids
            (B) TYPE: amino acid
            (C) STRANDEDNESS:
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: protein

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

        Ala Leu Gly Glu Gly Asp Lys Val Lys Cys Phe His Cys Gly Gly Gly
        1   5   10   15

        Leu Thr Asp Trp Lys Pro Ser Glu Asp Pro Trp Glu Gln His Ala Lys
                    20   25   30

        Trp Tyr Pro Gly Cys Lys Tyr Leu Leu Glu Gln Lys Gly Gln Glu Tyr
                    35   40   45

        Ile Asn
        50

(2) INFORMATION FOR SEQ ID NO:29:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 50 amino acids
            (B) TYPE: amino acid
            (C) STRANDEDNESS:
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: protein

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

        Ala Leu Gly Glu Gly Asp Lys Val Lys Cys Phe His Cys Gly Gly Gly
        1   5   10   15

        Leu Thr Asp Trp Lys Pro Ser Glu Asp Pro Trp Glu Gln His Ala Lys
                    20   25   30

        Trp Tyr Pro Gly Cys Lys Tyr Leu Leu Asp Glu Lys Gly Gln Glu Tyr
                    35   40   45

        Ile Asn
        50

(2) INFORMATION FOR SEQ ID NO:30:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 50 amino acids

          (B) TYPE: amino acid
          (C) STRANDEDNESS:
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: protein

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Tyr Val Gly Asn Ser Asp Asp Val Lys Cys Phe Cys Cys Asp Gly Gly
1   5   10   15

Leu Arg Cys Trp Glu Ser Gly Asp Asp Pro Trp Val Gln His Ala Lys
         20   25   30

Trp Phe Pro Arg Cys Glu Tyr Leu Ile Arg Ile Lys Gly Gln Glu Phe
     35   40   45

Ile Arg
    50

(2) INFORMATION FOR SEQ ID NO:31:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 50 amino acids
          (B) TYPE: amino acid
          (C) STRANDEDNESS:
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: protein

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Tyr Val Gly Arg Asn Asp Asp Val Lys Cys Phe Gly Cys Asp Gly Gly
1   5   10   15

Leu Arg Cys Trp Glu Ser Gly Asp Asp Pro Trp Val Glu His Ala Lys
         20   25   30

Trp Phe Pro Arg Cys Glu Phe Leu Ile Arg Met Lys Gly Gln Glu Phe
     35   40   45

Val Asp
    50

(2) INFORMATION FOR SEQ ID NO:32:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 50 amino acids
          (B) TYPE: amino acid
          (C) STRANDEDNESS:
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
Ala Leu Gly Glu Gly Asp Lys Val Lys Cys Phe His Cys Gly Gly Gly
1               5                   10              15

Leu Thr Asp Trp Lys Pro Ser Glu Asp Pro Trp Glu Gln His Ala Lys
            20              25          30

Trp Tyr Pro Gly Cys Lys Tyr Leu Leu Glu Gln Lys Gly Gln Glu Tyr
        35              40              45

Ile Asn
50
```

(2) INFORMATION FOR SEQ ID NO:33:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 50 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
Tyr Gln Lys Ile Gly Asp Gln Val Arg Cys Phe His Cys Asn Ile Gly
1               5                   10              15

Leu Arg Ser Trp Gln Lys Glu Asp Glu Pro Trp Phe Glu His Ala Lys
            20              25          30

Trp Ser Pro Lys Cys Gln Phe Val Leu Leu Ala Lys Gly Pro Ala Tyr
        35              40              45

Val Ser
50
```

(2) INFORMATION FOR SEQ ID NO:34:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 142 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS:
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
Met Gly Ala Pro Thr Leu Pro Pro Ala Trp Gln Pro Phe Leu Lys Asp
1               5                   10              15
```

```
His Arg Ile Ser Thr Phe Lys Asn Trp Pro Phe Leu Glu Gly Cys Ala
         20                  25                  30

Cys Thr Pro Glu Arg Met Ala Glu Ala Gly Phe Ile His Cys Pro Thr
         35                  40                  45

Glu Asn Glu Pro Asp Leu Ala Gln Cys Phe Phe Cys Phe Lys Glu Leu
         50                  55                  60

Glu Gly Trp Glu Pro Asp Asp Asp Pro Ile Glu Glu His Lys Lys His
         65             70                  75                  80

Ser Ser Gly Cys Ala Phe Leu Ser Val Lys Lys Gln Phe Glu Glu Leu
                     85                  90                  95

Thr Leu Gly Glu Phe Leu Lys Leu Asp Arg Glu Arg Ala Lys Asn Lys
         100                  105                  110

Ile Ala Lys Glu Thr Asn Asn Lys Lys Lys Glu Phe Glu Glu Thr Ala
         115                  120                  125

Lys Lys Val Arg Arg Ala Ile Glu Gln Leu Ala Ala Met Asp
         130                  135                  140
```

coiled a coil

(2) INFORMATION FOR SEQ ID NO:35:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 14796 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
TCTAGACATG CGGATATATT CAAGCTGGGC ACAGCACAGC AGCCCCACCC CAGGCAGCTT        60

GAAATCAGAG CTGGGGTCCA AAGGGACCAC ACCCCGAGGG ACTGTGTGGG GGTCGGGGCA       120

CACAGGCCAC TGCTTCCCCC CGTCTTTCTC AGCCATTCCT GAAGTCAGCC TCACTCTGCT       180

TCTCAGGGAT TTCAAATGTG CAGAGACTCT GGCACTTTTG TAGAAGCCCC TTCTGGTCCT       240

AACTTACACC TGGATGCTGT GGGGCTGCAG CTGCTGCTCG GGCTCGGGAG GATGCTGGGG       300

GCCCGGTGCC CATGAGCTTT TGAAGCTCCT GGAACTCGGT TTTGAGGGTG TTCAGGTCCA       360

GGTGGACACC TGGGCTGTCC TTGTCCATGC ATTTGATGAC ATTGTGTGCA GAAGTGAAAA       420

GGAGTTAGGC CGGGCATGCT GGCTTATGCC TGTAATCCCA GCACTTTGGG AGGCTGAGGC       480

GGGTGGATCA CGAGGTCAGG AGTTCAATAC CAGCCTGGCC AAGATGGTGA AACCCCGTCT       540

CTACTAAAAA TACAAAAAAA TTAGCCGGGC ATGGTGGCGG GCGCATGTAA TCCCAGCTAC       600

TGGGGGGGCT GAGGCAGAGA ATTGCTGGAA CCCAGGAGAT GGAGGTTGCA GTGAGCCAAG       660
```

```
ATTGTGCCAC TGCACTGCAC TCCAGCCTGG CGACAGAGCA AGACTCTGTC TCAAAAAAAA    720

AAAAAAAAAG TGAAAGGAG TTGTTCCTTT CCTCCCTCCT GAGGGCAGGC AACTGCTGCG     780

GTTGCCAGTG GAGGTGGTGC GTCCTTGGTC TGTGCCTGGG GGCCACCCCA GCAGAGGCCA    840

TGGTGGTGCC AGGGCCCGGT TAGCGAGCCA ATCAGCAGGA CCCAGGGGCG ACCTGCCAAA    900

GTCAACTGGA TTTGATAACT GCAGCGAAGT TAAGTTTCCT GATTTTGATG ATTGTGTTGT    960

GGTTGTGTAA GAGAATGAAG TATTTCGGGG TAGTATGGTA ATGCCTTCAA CTTACAAACG   1020

GTTCAGGTAA ACCACCCATA TACATACATA TACATGCATG TGATATATAC ACATACAGGG   1080

ATGTGTGTGT GTTCACATAT ATGAGGGGAG AGAGACTAGG GGAGAGAAAG TAGGTTGGGG   1140

AGAGGGAGAG AGAAAGGAAA ACAGGAGACA GAGAGAGAGC GGGGAGTAGA GAGAGGGAAG   1200

GGGTAAGAGA GGGAGAGGAG GAGAGAAAGG GAGGAAGAAG CAGAGAGTGA ATGTTAAAGG   1260

AAACAGGCAA AACATAAACA GAAAATCTGG GTGAAGGGTA TATGAGTATT CTTTGTACTA   1320

TTCTTGCAAT TATCTTTTAT TTAAATTGAC ATCGGGCCGG GCGCAGTGGC TCACATCTGT   1380

AATCCCAGCA CTTTGGGAGG CCGAGGCAGG CAGATCACTT GAGGTCAGGA GTTTGAGACC   1440

AGCCTGGCAA ACATGGTGAA ACCCCATCTC TACTAAAAAT ACAAAAATTA GCCTGGTGTG   1500

GTGGTGCATG CCTTTAATCT CAGCTACTCG GGAGGCTGAG GCAGGAGAAT CGCTTGAACC   1560

CGTGGCGGGG AGGAGGTTGC AGTGAGCTGA GATCATGCCA CTGCACTCCA GCCTGGGCGA   1620

TAGAGCGAGA CTCAGTTTCA AATAAATAAA TAAACATCAA AATAAAAGT TACTGTATTA    1680

AAGAATGGGG GCGGGGTGGG AGGGGTGGGG AGAGGTTGCA AAAATAAATA AATAAATAAA   1740

TAAACCCCAA AATGAAAAAG ACAGTGGAGG CACCAGGCCT GCGTGGGGCT GGAGGGCTAA   1800

TAAGGCCAGG CCTCTTATCT CTGGCCATAG AACCAGAGAA GTGAGTGGAT GTGATGCCCA   1860

GCTCCAGAAG TGACTCCAGA ACACCCTGTT CCAAAGCAGA GGACACACTG ATTTTTTTTT   1920

TAATAGGCTG CAGGACTTAC TGTTGGTGGG ACGCCCTGCT TTGCGAAGGG AAAGGAGGAG   1980

TTTGCCCTGA GCACAGGCCC CCACCCTCCA CTGGGCTTTC CCCAGCTCCC TTGTCTTCTT   2040

ATCACGGTAG TGGCCCAGTC CCTGGCCCCT GACTCCAGAA GGTGGCCCTC CTGGAAACCC   2100

AGGTCGTGCA GTCAACGATG TACTCGCCGG GACAGCGATG TCTGCTGCAC TCCATCCCTC   2160

CCCTGTTCAT TTGTCCTTCA TGCCCGTCTG GAGTAGATGC TTTTTGCAGA GGTGGCACCC   2220

TGTAAAGCTC TCCTGTCTGA CTTTTTTTTT TTTTTAGAC TGAGTTTTGC TCTTGTTGCC    2280

TAGGCTGGAG TGCAATGGCA CAATCTCAGC TCACTGCACC CTCTGCCTCC CGGGTTCAAG   2340

CGATTCTCCT GCCTCAGCCT CCCGAGTAGT TGGGATTACA GGCATGCACC ACCACGCCCA   2400

GCTAATTTTT GTATTTTTAG TAGAGACAAG GTTTCACCGT GATGGCCAGG CTGGTCTTGA   2460

ACTCCAGGAC TCAAGTGATG CTCCTGCCTA GGCCTCTCAA AGTGTTGGGA TTACAGGCGT   2520
```

```
GAGCCACTGC ACCCGGCCTG CACGCGTTCT TTGAAAGCAG TCGAGGGGGC GCTAGGTGTG    2580

GGCAGGGACG AGCTGGCGCG GCGTCGCTGG GTGCACCGCG ACCACGGGCA GAGCCACGCG    2640

GCGGGAGGAC TACAACTCCC GGCACACCCC GCGCCGCCCC GCCTCTACTC CCAGAAGGCC    2700

GCGGGGGGTG GACCGCCTAA GAGGGCGTGC GCTCCCGACA TGCCCCGCGG CGCGCCATTA    2760

ACCGCCAGAT TTGAATCGCG GGACCCGTTG GCAGAGGTGG CGGCGGCGGC ATGGGTGCCC    2820

CGACGTTGCC CCCTGCCTGG CAGCCCTTTC TCAAGGACCA CCGCATCTCT ACATTCAAGA    2880

ACTGGCCCTT CTTGGAGGGC TGCGCCTGCA CCCCGGAGCG GGTGAGACTG CCCGGCCTCC    2940

TGGGGTCCCC CACGCCCGCC TTGCCCTGTC CCTAGCGAGG CCACTGTGAC TGGGCCTCGG    3000

GGGTACAAGC CGCCCTCCCC TCCCCGTCCT GTCCCCAGCG AGGCCACTGT GGCTGGGCCC    3060

CTTGGGTCCA GGCCGGCCTC CCCTCCCTGC TTTGTCCCCA TCGAGGCCTT TGTGGCTGGG    3120

CCTCGGGGTT CCGGGCTGCC ACGTCCACTC ACGAGCTGTG CTGTCCCTTG CAGATGGCCG    3180

AGGCTGGCTT CATCCACTGC CCCACTGAGA ACGAGCCAGA CTTGGCCCAG TGTTTCTTCT    3240

GCTTCAAGGA GCTGGAAGGC TGGGAGCCAG ATGACGACCC CATGTAAGTC TTCTCTGGCC    3300

AGCCTCGATG GGCTTTGTTT TGAACTGAGT TGTCAAAAGA TTTGAGTTGC AAAGACACTT    3360

AGTATGGGAG GGTTGCTTTC CACCCTCATT GCTTCTTAAA CAGCTGTTGT GAACGGATAC    3420

CTCTCTATAT GCTGGTGCCT TGGTGATGCT TACAACCTAA TTAAATCTCA TTTGACCAAA    3480

ATGCCTTGGG GTGGACGTAA GATGCCTGAT GCCTTTCATG TTCAACAGAA TACATCAGCA    3540

GACCCTGTTG TTGTGAACTC CCAGGAATGT CCAAGTGCTT TTTTTGAGAT TTTTTAAAAA    3600

ACAGTTTAAT TGAAATATAA CCTACACAGC ACAAAAATTA CCCTTTGAAA GTGTGCACTT    3660

CACACTTTCG GAGGCTGAGG CGGGCGGATC ACCTGAGGTC AGGAGTTCAA GACCTGCCTG    3720

GCCAACTTGG CGAAACCCCG TCTCTACTAA AAATACAAAA ATTAGCCGGG CATGGTAGCG    3780

CACGCCCGTA ATCCCAGCTA CTCGGGAGGC TAAGGCAGGA GAATCGCTTG AACCTGGGAG    3840

GCGGAGGTTG CAGTGAGCCG AGATTGTGCC AATGCACTCC AGCCTCGGCG ACAGAGCGAG    3900

ACTCCGTCAT AAAAATAAAA AATTGAAAAA AAAAAAGAA AGAAAGCATA TACTTCAGTG     3960

TTGTTCTGGA TTTTTTTCTT CAAGATGCCT AGTTAATGAC AATGAAATTC TGTACTCGGA    4020

TGGTATCTGT CTTTCCACAC TGTAATGCCA TATTCTTTTC TCACCTTTTT TTCTGTCGGA    4080

TTCAGTTGCT TCCACAGCTT TAATTTTTTT CCCCTGGAGA ATCACCCCAG TTGTTTTTCT    4140

TTTTGGCCAG AAGAGAGTAG CTGTTTTTTT TCTTAGTATG TTTGCTATGG TGGTTATACT    4200

GCATCCCCGT AATCACTGGG AAAAGATCAG TGGTATTCTT CTTGAAAATG AATAAGTGTT    4260

ATGATATTTT CAGATTAGAG TTACAACTGG CTGTCTTTTT GGACTTGTG TGGCCATGTT     4320
```

```
TTCATTGTAA TGCAGTTCTG GTAACGGTGA TAGTCAGTTA TACAGGGAGA CTCCCCTAGC    4380

AGAAAATGAG AGTGTGAGCT AGGGGGTCCC TTGGGGAACC CGGGGCAATA ATGCCCTTCT    4440

CTGCCCTTAA TCCTTACAGT GGGCCGGGCA CGGTGGCTTA CGCCTGTAAT ACCAGCACTT    4500

TGGGAGGCCG AGGCGGGCGG ATCACGAGGT CAGGAGATCG AGACCATCTT GGCTAATACG    4560

GTGAAACCCC GTCTCCACTA AAAATACAAA AAATTAGCCG GGCGTGGTGG TGGGCGCCTG    4620

TAGTCCCAGC TACTCGGGAG GCTGAGGCAG GAGAATGGCG TGAACCCAGG AGGCGGAGCT    4680

TGCAGTGAGC CGAGATTGCA CCACTGCACT CCAGCCTGGG CGACAGAATG AGACTCCGTC    4740

TCAAAAAAAA AAAAAAAGA AAAAAATCTT TACAGTGGAT TACATAACAA TTCCAGTGAA    4800

ATGAAATTAC TTCAAACAGT TCCTTGAGAA TGTTGGAGGG ATTTGACATG TAATTCCTTT    4860

GGACATATAC CATGTAACAC TTTTCCAACT AATTGCTAAG GAAGTCCAGA TAAAATAGAT    4920

ACATTAGCCA CACAGATGTG GGGGGAGATG TCCACAGGGA GAGAGAAGGT GCTAAGAGGT    4980

GCCATATGGG AATGTGGCTT GGGCAAAGCA CTGATGCCAT CAACTTCAGA CTTGACGTCT    5040

TACTCCTGAG GCAGAGCAGG GTGTGCCTGT GGAGGGCGTG GGGAGGTGGC CCGTGGGGAG    5100

TGGACTGCCG CTTTAATCCC TTCAGCTGCC TTTCCGCTGT TGTTTTGATT TTTCTAGAGA    5160

GGAACATAAA AAGCATTCGT CCGGTTGCGC TTTCCTTTCT GTCAAGAAGC AGTTTGAAGA    5220

ATTAACCCTT GGTGAATTTT TGAAACTGGA CAGAGAAAGA GCCAAGAACA AAATTGTATG    5280

TATTGGGAAT AAGAACTGCT CAAACCCTGT TCAATGTCTT TAGCACTAAA CTACCTAGTC    5340

CCTCAAAGGG ACTCTGTGTT TTCCTCAGGA AGCATTTTTT TTTTTTTCT GAGATAGAGT    5400

TTCACTCTTG TTGCCCAGGC TGGAGTGCAA TGGTGCAATC TTGGCTCACT GCAACCTCTG    5460

CCTCTCGGGT TCAAGTGATT CTCCTGCCTC AGCCTCCCAA GTAACTGGGA TTACAGGGAA    5520

GTGCCACCAC ACCCAGCTAA TTTTTGTATT TTTAGTAGAG ATGGGGTTTC ACCACATTGC    5580

CCAGGCTGGT CTTGAACTCC TGACCTCGTG ATTCGCCCAC CTTGGCCTCC CAAAGTGCTG    5640

GGATTACAGG CGTGAACCAC CACGCCTGGC TTTTTTTTT TTGTTCTGAG ACACAGTTTC    5700

ACTCTGTTAC CCAGGCTGGA GTAGGGTGGC CTGATCTCGG ATCACTGCAA CCTCCGCCTC    5760

CTGGGCTCAA GTGATTTGCC TGCTTCAGCC TCCCAAGTAG CCGAGATTAC AGGCATGTGC    5820

CACCACACCC AGGTAATTTT TGTATTTTTG GTAGAGACGA GGTTTCACCA TGTTGGCCAG    5880

GCTGGTTTTG AACTCCTGAC CTCAGGTGAT CCACCCGCCT CAGCCTCCCA AAGTGCTGAG    5940

ATTATAGGTG TGAGCCACCA CACCTGGCCT CAGGAAGTAT TTTTATTTTT AAATTTATTT    6000

ATTTATTTGA GATGGAGTCT TGCTCTGTCG CCCAGGCTAG AGTGCAGCGA CGGGATCTCG    6060

GCTCACTGCA AGCTCCGCCC CCCAGGTTCA AGCCATTCTC CTGCCTCAGC CTCCCGAGTA    6120

GCTGGGACTA CAGGCGCCCG CCACCACACC CGGCTAATTT TTTTGTATTT TTAGTAGAGA    6180
```

```
CGGGTTTTCA CCGTGTTAGC CAGGAGGGTC TTGATCTCCT GACCTCGTGA TCTGCCTGCC    6240

TCGGCCTCCC AAAGTGCTGG GATTACAGGT GTGAGCCACC ACACCCGGCT ATTTTTATTT    6300

TTTTGAGACA GGGACTCACT CTGTCACCTG GGCTGCAGTG CAGTGGTACA CCATAGCTCA    6360

CTGCAGCCTC GAACTCCTGA GCTCAAGTGA TCCTCCCACC TCATCCTCAC AAGTAATTGG    6420

GACTACAGGT GCACCCACC ATGCCCACCT AATTTATTTA TTTATTTATT TATTTATTTT      6480

CATAGAGATG AGGGTTCCCT GTGTTGTCCA GGCTGGTCTT GAACTCCTGA GCTCACGGGA    6540

TCCTTTTGCC TGGGCCTCCC AAAGTGCTGA GATTACAGGC ATGAGCCACC GTGCCCAGCT    6600

AGGAATCATT TTTAAAGCCC CTAGGATGTC TGTGTGATTT TAAAGCTCCT GGAGTGTGGC    6660

CGGTATAAGT ATATACCGGT ATAAGTAAAT CCCACATTTT GTGTCAGTAT TTACTAGAAA    6720

CTTAGTCATT TATCTGAAGT TGAAATGTAA CTGGGCTTTA TTTATTTATT TATTTATTTA    6780

TTTATTTTTA ATTTTTTTTT TTGAGACGAG TCTCACTTTG TCACCCAGGC TGGAGTGCAG    6840

TGGCACGATC TCGGCTCACT GCAACCTCTG CCTCCCGGGG TCAAGCGATT CTCCTGCCTT    6900

AGCCTCCCGA GTAGCTGGGA CTACAGGCAC GCACCACCAT GCCTGGCTAA TTTTTGTATT    6960

TTTAGTAGAC GGGGTTTCAC CATGCTGGCC AAGCTGGTCT CAAACTCCTG ACCTTGTGAT    7020

CTGCCCGCTT TAGCCTCCCA GAGTGCTGGG ATTACAGGCA TGAGCCACCA TGCGTGGTCT    7080

TTTTAAAATT TTTTGATTTT TTTTTTTTTT GAGACAGAGC CTTGCTCTGT CGCCCAGGCT    7140

GGAGTGCAGT GGCACGATCT CAGCTCACTA CAAGCTCCGC CTCCCGGGTT CACGCCATTC    7200

TTCTGCCTCA GCCTCCTGAG TAGCTGGGAC TACAGGTGCC CACCACCACG CCTGGCTAAT    7260

TTTTTTTGGT ATTTTTATTA GAGACAAGGT TTCATCATGT TGGCCAGGCT GGTCTCAAAC    7320

TCCTGACCTC AAGTGATCTG CCTGCCTCGG CCTCCCAAAG CGCTGAGATT ACAGGTGTGA    7380

TCTACTGCGC CAGGCCTGGG CGTCATATAT TCTTATTTGC TAAGTCTGGC AGCCCCACAC    7440

AGAATAAGTA CTGGGGGATT CCATATCCTT GTAGCAAAGC CCTGGGTGGA GAGTCAGGAG    7500

ATGTTGTAGT TCTGTCTCTG CCACTTGCAG ACTTTGAGTT TAAGCCAGTC GTGCTCATGC    7560

TTTCCTTGCT AAATAGAGGT TAGACCCCCT ATCCCATGGT TTCTCAGGTT GCTTTTCAGC    7620

TTGAAAATTG TATTCCTTTG TAGAGATCAG CGTAAAATAA TTCTGTCCTT ATATGTGGCT    7680

TTATTTTAAT TTGAGACAGA GTGTCACTCA GTCGCCCAGG CTGGAGTGTG GTGGTGCGAT    7740

CTTGGCTCAC TGCGACCTCC ACCTCCCAGG TTCAAGCGAT TCTCGTGCCT CAGGCTCCCA    7800

AGTAGCTGAG ATTATAGGTG TGTGCCACCA GGCCCAGCTA ACTTTGTAT TTTTAGTAGA     7860

GACAGGGTTT TGCCATGTTG GCTAAGCTGG TCTCGAACTC CTGGCCTCAA GTGATCTGCC    7920

CGCCTTGGCA TCCCAAAGTG CTGGGATTAC AGGTGTGAAC CACCACACCT GGCCTCAATA    7980
```

```
TAGTGGCTTT TAAGTGCTAA GGACTGAGAT TGTGTTTTGT CAGGAAGAGG CCAGTTGTGG    3040

GTGAAGCATG CTGTGAGAGA GCTTGTCACC TGGTTGAGGT TGTGGGAGCT GCAGCGTGGG    3100

AACTGGAAAG TGGGCTGGGG ATCATCTTTT TCCAGGTCAG GGGTCAGCCA GCTTTTCTGC    8160

AGCGTGCCAT AGACCATCTC TTAGCCCTCG TGGGTCAGAG TCTCTGTTGC ATATTGTCTT    8220

TTGTTGTTTT TCACAACCTT TTAGAAACAT AAAAAGCATT CTTAGCCCGT GGGCTGGACA    8280

AAAAAAGGCC ATGACGGGCT GTATGGATTT GGCCCAGCAG GCCCTTGCTT GCCAAGCCCT    8340

GTTTTAGACA AGGAGCAGCT TGTGTGCCTG GAACCATCAT GGGCACAGGG GAGGAGCAGA    8400

GTGGATGTGG AGGTGTGAGC TGGAAACCAG GTCCCAGAGC GCTGAGAAAG ACAGAGGGTT    8460

TTTGCCCTTG CAAGTAGAGC AACTGAAATC TGACACCATC CAGTTCCAGA AAGCCCTGAA   ·8520

GTGCTGGTGG ACGCTGCGGG GTGCTCCGCT CTAGGGTTAC AGGGATGAAG ATGCAGTCTG    8580

GTAGGGGGAG TCCACTCACC TGTTGGAAGA TGTGATTAAG AAAAGTAGAC TTTCAGGGCC    8640

GGGCATGGTG GCTCACGCCT GTAATCCCAG CACTTTGGGA GGCCGAGGCG GGTGGATCAC    8700

GAGGTCAGGA GATCGAGACC ATCCTGGCTA ACATGGTGAA ACCCCGTCTT TACTAAAAAT    8760

ACAAAAAATT AGCTGGGCGT GGTGGCGGGC GCCTGTAGTC CCAGCTACTC GGGAGGCTGA    8820

GGCAGGAGAA TGGCGTGAAC CTGGGAGGTG GAGCTTGCTG TGAGCCGAGA TCGCGCCACT    8880

GCACTCCAGC CTGGGCGACA GAGCGAGACT CCGTCTCAAA AAAAAAAAA AAAGTAGGCT    8940

TTCATGATGT GTGAGCTGAA GGCGCAGTAG GCAGAAGTAG AGGCCTCAGT CCCTGCAGGA    9000

GACCCCTCGG TCTCTATCTC CTGATAGTCA GACCCAGCCA CACTGGAAAG AGGGGAGACA    9060

TTACAGCCTG CGAGAAAGT AGGGAGATTT AAAAACTGCT TGGCTTTTAT TTTGAACTGT    9120

TTTTTTTGTT TGTTTGTTTT CCCCAATTCA GAATACAGAA TACTTTATG GATTTGTTTT    9180

TATTACTTTA ATTTTGAAAC AATATAATCT TTTTTTTGTT GTTTTTTTGA GACAGGGTCT    9240

TACTCTGTCA CCCAGGCTGA GTGCAGTGGT GTGATCTTGG CTCACCTCAG CCTCGACCCC    9300

CTGGGCTCAA ATGATTCTCC CACCTCAGCT TCCCAAGTAG CTGGGACCAC AGGTGCGTGT    9360

GTTGCGCTAT ACAAATCCTG AAGACAAGGA TGCTGTTGCT GGTGATGCTG GGGATTCCCA    9420

AGATCCCAGA TTTGATGGCA GGATGCCCCT GTCTGCTGCC TTGCCAGGGT GCCAGGAGGG    9480

CGCTGCTGTG GAAGCTGAGG CCCGGCCATC CAGGGCGATG CATTGGGCGC TGATTCTTGT    9540

TCCTGCTGCT GCCTCGGTGC TTAGCTTTTG AAACAATGAA ATAAATTAGA ACCAGTGTGA    9600

AAATCGATCA GGGAATAAAT TTAATGTGGA AATAAACTGA ACAACTTAGT TCTTCATAAG    9660

AGTTTACTTG GTAAATACTT GTGATGAGGA CAAAACGAAG CACTAGAAGG AGAGGCGAGT    9720

TGTAGACCTG GGTGGCAGGA GTGTTTTGTT TGTTTTCTTT GGCAGGGTCT TGCTCTGTTG    9780

CTCAGGCTGG AGTACAGTGG CACAATCACA GCTCACTATA GCCTCGACCT CCTGGACTCA    9840
```

```
AGCAATCCTC CTGCCTCAGC CTCCCAGTAG CTGGGACTAC AGGCGCATGC CACCATGCCT      9900

GGCTAATTTT AAATTTTTTT TTTTCTCTTT TTTGAGATGG AATCTCACTC TGTCGCCCAG      9960

GCTGGAGTGC AGTGGCGTGA TCTCGGCTGA CGGCAAGCTC CGCCTCCCAG GTTCACTCCA     10020

TTCGCCTGCC TCAGCCTCCC AAGTAGCTGG GACTACAGGC GCTGGGATTA CAAACCCAAA     10080

CCCAAAGTGC TGGGATTACA GGCGTGAGCC ACTGCACCCG GCCTGTTTTG TCTTTCAATA     10140

GCAAGAGTTG TGTTTGCTTC GCCCCTACCT TTAGTGGAAA AATGTATAAA ATGGAGATAT     10200

TGACCTCCAC ATTGGGGTGG TTAAATTATA GCATGTATGC AAAGGAGCTT CGCTAATTTA     10260

AGGCTTTTTT GAAAGAGAAG AAACTGAATA ATCCATGTGT GTATATATAT TTTAAAAGCC     10320

ATGGTCATCT TTCCATATCA GTAAAGCTGA GGCTCCCTGG GACTGCAGAG TTGTCCATCA     10380

CAGTCCATTA TAAGTGCGCT GCTGGGCCAG GTGCAGTGGC TTGTGCCTGA ATCCCAGCAC     10440

TTTGGGAGGC CAAGGCAGGA GGATTCATTG AGCCCAGGAG TTTTGAGGCG AGCCTGGGCA     10500

ATGTGGCCAG ACCTCATCTC TTCAAAAAAT ACACAAAAAA TTAGCCAGGC ATGGTGGCAC     10560

GTGCCTGTAG TCTCAGCTAC TCAGGAGGCT GAGGTGGGAG GATCACTTTG AGCCTTGCAG     10620

GTCAAAGCTG CAGTAAGCCA TGATCTTGCC ACTGCATTCC AGCCTGGATG ACAGAGCGAG     10680

ACCCTGTCTC TAAAAAAAAA AAAAACCAAA CGGTGCACTG TTTTCTTTTT TCTTATCAAT     10740

TTATTATTTT TAAATTAAAT TTTCTTTTAA TAATTTATAA ATTATAAATT TATATTAAAA     10800

AATGACAAAT TTTTATTACT TATACATGAG GTAAAACTTA GGATATATAA AGTACATATT     10860

GAAAAGTAAT TTTTTGGCTG GCACAGTGGC TCACACCTGT AATCCCAGCA CTTTGGGAGG     10920

CCGTGGCGGG CAGATCACAT GAGATCATGA GTTCGAGACC AACCTGACCA ACATGGAGAG     10980

ACCCCATCTC TACTAAAAAT ACAAAATTAG CCGGGGTGGT GGCGCATGCC TGTAATCCCA     11040

GCTACTCGGG AGGCTGAGGC AGGAGAATCT CTTGAACCCG GGAGGCAGAG GTTGCGGTGA     11100

GCCAAGATCG TGCCTTTGCA CACCAGCCTA GGCAACAAGA GCGAAAGTCC GTCTCAAAAA     11160

AAAAGTAATT TTTTTAAGT TAACCTCTGT CAGCAAACAA ATTTAACCCA ATAAAGGTCT     11220

TTGTTTTTTA ATGTAGTAGA GGAGTTAGGG TTTATAAAAA ATATGGTAGG GAAGGGGGTC     11280

CCTGGATTTG CTAATGTGAT TGTCATTTGC CCCTTAGGAG AGAGCTCTGT TAGCAGAATG     11340

AAAAAATTGG AAGCCAGATT CAGGGAGGGA CTGGAAGCAA AAGAATTTCT GTTCGAGGAA     11400

GAGCCTGATG TTTGCCAGGG TCTGTTTAAC TGGACATGAA GAGGAAGGCT CTGGACTTTC     11460

CTCCAGGAGT TTCAGGAGAA AGGTAGGGCA GTGGTTAAGA GCAGAGCTCT GCCTAGACTA     11520

GCTGGGGTGC CTAGACTAGC TGGGGTGCCC AGACTAGCTG GGGTGCCTAG ACTAGCTGGG     11580

TACTTTGAGT GGCTCCTTCA GCCTGGACCT CGGTTTCCTC ACCTGTATAG TAGAGATATG     11640
```

```
GGAGCACCCA GCGCAGGATC ACTGTGAACA TAAATCAGTT AATGGAGGAA GCAGGTAGAG   11700

TGGTGCTGGG TGCATACCAA GCACTCCGTC AGTGTTTCCT GTTATTCGAT GATTAGGAGG   11760

CAGCTTAAAC TAGAGGGAGT TGAGCTGAAT CAGGATGTTT GTCCCAGGTA GCTGGGAATC   11820

TGCCTAGCCC AGTGCCCAGT TTATTTAGGT GCTCTCTCAG TGTTCCCTGA TTGTTTTTTC   11880

CTTTGTCATC TTATCTACAG GATGTGACTG GGAAGCTCTG GTTTCAGTGT CATGTGTCTA   11940

TTCTTTATTT CCAGGCAAAG GAAACCAACA ATAAGAAGAA AGAATTTGAG GAAACTGCGA   12000

AGAAAGTGCG CCGTGCCATC GAGCAGCTGG CTGCCATGGA TTGAGGCCTC TGGCCGGAGC   12060

TGCCTGGTCC CAGAGTGGCT GCACCACTTC CAGGGTTTAT TCCCTGGTGC CACCAGCCTT   12120

CCTGTGGGCC CCTTAGCAAT GTCTTAGGAA AGGAGATCAA CATTTTCAAA TTAGATGTTT   12180

CAACTGTGCT CCTGTTTTGT CTTGAAAGTG GCACCAGAGG TGCTTCTGCC TGTGCAGCGG   12240

GTGCTGCTGG TAACAGTGGC TGCTTCTCTC TCTCTCTCTC TTTTTGGGG GCTCATTTTT   12300

GCTGTTTTGA TTCCCGGGCT TACCAGGTGA GAAGTGAGGG AGGAAGAAGG CAGTGTCCCT   12360

TTTGCTAGAG CTGACAGCTT TGTTCGCGTG GGCAGAGCCT TCCACAGTGA ATGTGTCTGG   12420

·ACCTCATGTT GTTGAGGCTG TCACAGTCCT GAGTGTGGAC TTGGCAGGTG CCTGTTGAAT   12480

CTGAGCTGCA GGTTCCTTAT CTGTCACACC TGTGCCTCCT CAGAGGACAG TTTTTTTGTT   12540

GTTGTGTTTT TTTGTTTTTT TTTTTGGTA GATGCATGAC TTGTGTGTGA TGAGAGAATG   12600

GAGACAGAGT CCCTGGCTCC TCTACTGTTT AACAACATGG CTTTCTTATT TTGTTTGAAT   12660

TGTTAATTCA CAGAATAGCA CAAACTACAA TTAAAACTAA GCACAAAGCC ATTCTAAGTC   12720

ATTGGGGAAA CGGGGTGAAC TTCAGGTGGA TGAGGAGACA GAATAGAGTG ATAGGAAGCG   12780

TCTGGCAGAT ACTCCTTTTG CCACTGCTGT GTGATTAGAC AGGCCCAGTG AGCCGCGGGG   12840

CACATGCTGG CCGCTCCTCC CTCAGAAAAA GGCAGTGGCC TAAATCCTTT TTAAATGACT   12900

TGGCTCGATG CTGTGGGGGA CTGGCTGGGC TGCTGCAGGC CGTGTGTCTG TCAGCCCAAC   12960

CTTCACATCT GTCACGTTCT CCACACGGGG GAGAGACGCA GTCCGCCCAG GTCCCCGCTT   13020

TCTTTGGAGG CAGCAGCTCC CGCAGGGCTG AAGTCTGGCG TAAGATGATG GATTTGATTC   13080

GCCCTCCTCC CTGTCATAGA GCTGCAGGGT GGATTGTTAC AGCTTCGCTG GAAACCTCTG   13140

GAGGTCATCT CGGCTGTTCC TGAGAAATAA AAAGCCTGTC ATTTCAAACA CTGCTGTGGA   13200

CCCTACTGGG TTTTTAAAAT ATTGTCAGTT TTTCATCGTC GTCCCTAGCC TGCCAACAGC   13260

CATCTGCCCA GACAGCCGCA GTGAGGATGA GCGTCCTGGC AGAGACGCAG TTGTCTCTGG   13320

GCGCTTGCCA GAGCCACGAA CCCCAGACCT GTTTGTATCA TCCGGGCTCC TTCCGGGCAG   13380

AAACAACTGA AAATGCACTT CAGACCCACT TATTTATGCC ACATCTGAGT CGGCCTGAGA   13440

TAGACTTTTC CCTCTAAACT GGGAGAATAT CACAGTGGTT TTTGTTAGCA GAAAATGCAC   13500
```
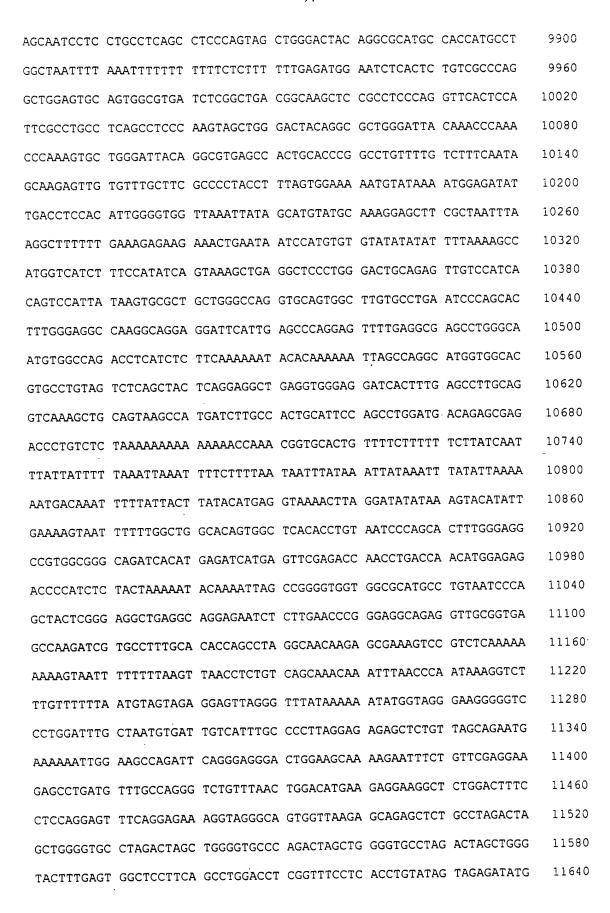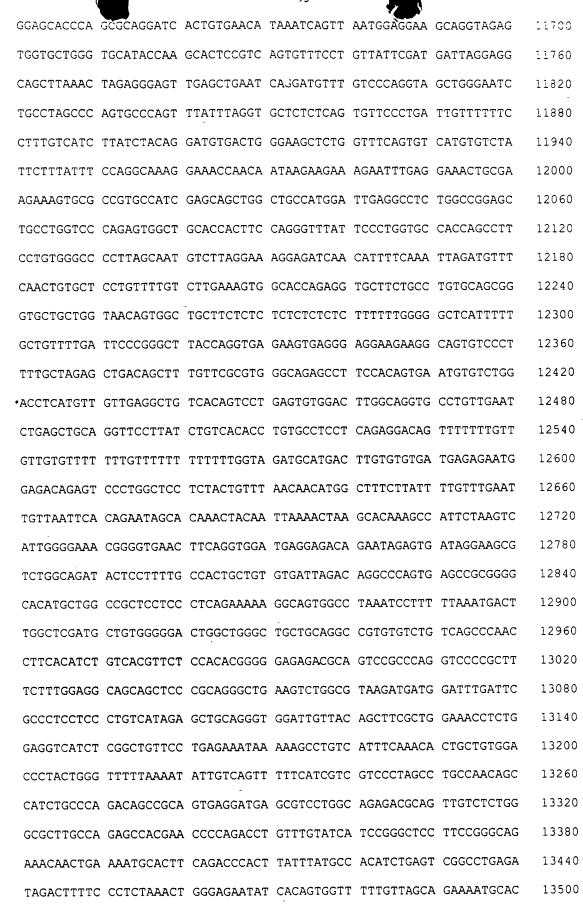
```
TCCAGCCTCT GTACTCATCT AAGCTGCTTA TTTTTGATAT TTGTGTCAGT CTGTAAATGG    13560
ATACTTCACT TTAATAACTG TTGCTTAGTA ATTGGCTTTG TAGAGAAGCT GGAAAAAAAT    13620
GGTTTTGTCT TCAACTCCTT TGCATGCCAG GCGGTGATGT GGATCTCGGC TTCTGTGAGC    13680
CTGTGCTGTG GGCAGGGCTG AGCTGGAGCC GCCCCTCTCA GCCCGCCTGC CACGGCCTTT    13740
CCTTAAAGGC CATCCTTAAA ACCAGACCCT CATGGCTGCC AGCACCTGAA AGCTTCCTCG    13800
ACATCTGTTA ATAAAGCCGT AGGCCCTTGT CTAAGCGCAA CCGCCTAGAC TTTCTTTCAG    13860
ATACATGTCC ACATGTCCAT TTTTCAGGTT CTCTAAGTTG GAGTGGAGTC TGGGAAGGGT    13920
TGTGAATGAG GCTTCTGGGC TATGGGTGAG GTTCCAATGG CAGGTTAGAG CCCCTCGGGC    13980
CAACTGCCAT CCTGGAAAGT AGAGACAGCA GTGCCCGCTG CCCAGAAGAG ACCAGCAAGC    14040
CAAACTGGAG CCCCCATTGC AGGCTGTCGC CATGTGGAAA GAGTAACTCA CAATTGCCAA    14100
TAAAGTCTCA TGTGGTTTTA TCTACTTTTT TTTTCTTTTT CTTTTTTTTT GAGACAAGGC    14160
CTTGCCCTCC CAGGCTGGAG TGCAGTGGAA TGACCACAGC TCACCGCAAC CTCAAATTCT    14220
TGCGTTCAAG TGAACCTCCC ACTTTAGCCT CCCAAGTAGC TGGGACTACA GGCGCACGCC    14280
ATCACACCCG GCTAATTGAA AAATTTTTT TTTTGTTTAG ATGGAATCTC ACTTTGTTGC    14340
CCAGGCTGGT CTCAAACTCC TGGGCTCAAG TGATCATCCT GCTTCAGCGT CCGACTTGTT    14400
GGTATTATAG GCGTGAGCCA CTGGGCCTGA CCTAGCTACC ATTTTTTAAT GCAGAAATGA    14460
AGACTTGTAG AAATGAAATA ACTTGTCCAG GATAGTCGAA TAAGTAACTT TTAGAGCTGG    14520
GATTTGAACC CAGGCAATCT GGCTCCAGAG CTGGGCCCTC ACTGCTGAAG GACACTGTCA    14580
GCTTGGGAGG GTGGCTATGG TCGGCTGTCT GATTCTAGGG AGTGAGGGCT GTCTTTAAAG    14640
CACCCCATTC CATTTTCAGA CAGCTTTGTC AGAAAGGCTG TCATATGGAG CTGACACCTG    14700
CCTCCCCAAG GCTTCCATAG ATCCTCTCTG TACATTGTAA CCTTTTATTT TGAAATGAAA    14760
                                                                    14796
ATTCACAGGA AGTTGTAAGG CTAGTACAGG GGATCC
```